

Biostat_projet

Claire et Lucie

01/12/2021

Le Danemark est un pays scandinave. Pour ce pays, nous allons étudier les différentes causes de mortalités en fonction de l'âge des individus, de la période où elles sont mortes.

Explications du jeu de données

```
mort <- read.csv("mortDK.csv", header = TRUE, sep = ";")
mort <- mort %>% rename("taux.deces" = rt,
  "Infections" = r1,
  "Cancer" = r2,
  "Tumeurs.Bénignes" = r3,
  "Endocrine" = r4,
  "Maladies.Sang" = r5,
  "Systeme.Nerveux" = r6,
  "Cerebrovasculaire" = r7,
  "Cardiaque" = r8,
  "Maladies.Respiratoires" = r9,
  "Maladies.Foie" = r10,
  "Maladies.Digestives" = r11,
  "Maladie.Génito_urinaire" = r12,
  "Non.Definis" = r13,
  "Mort.Naturelle" = r14,
  "Mort.Violente" = r15)

mort$taux.deces <- mort$taux.deces %>% as.numeric()
mort <- mort %>% filter(taux.deces <= 100) # Certains taux de décès sont supérieurs à 100
mort$sex <- mort$sex %>% factor( labels = c("H", "F"))
mort$periode <- mort$periode %>% factor( labels = c("43-47", "48-52",
  "53-57", "58-62",
  "63-67", "68-72",
  "73-77", "78-82",
  "83-87", "88-92"))
```

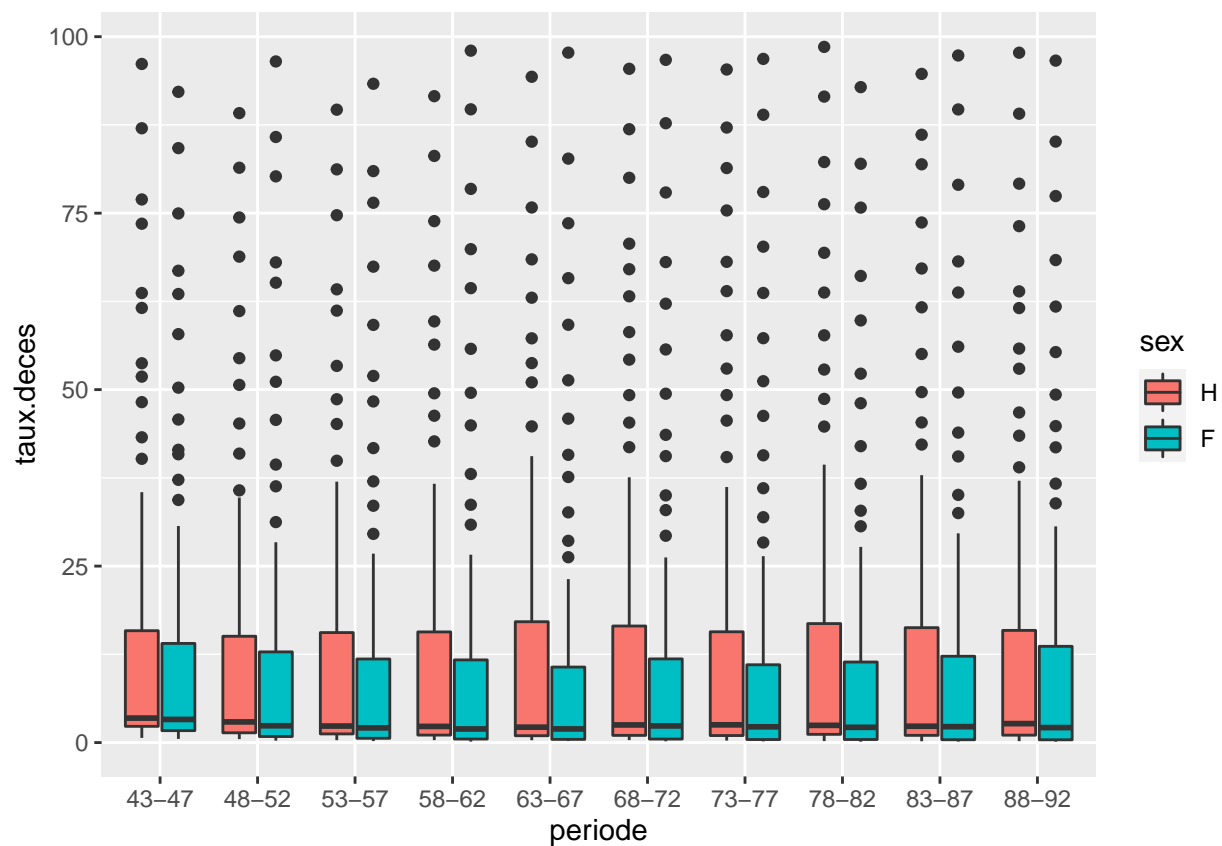
Le jeu de données est composé de 1820 observations et 21 variables. Ces variables sont l'âge, le sexe, la période, le nombre de décès, le taux de décès et 15 variables sur l'identification du décès. Lorsque qu'on parle d'identification du décès, on parle si la personne est morte d'un cancer, d'une mort naturelle, par exemple.

La variable *taux.deces* est calculée d'après la formule suivante :

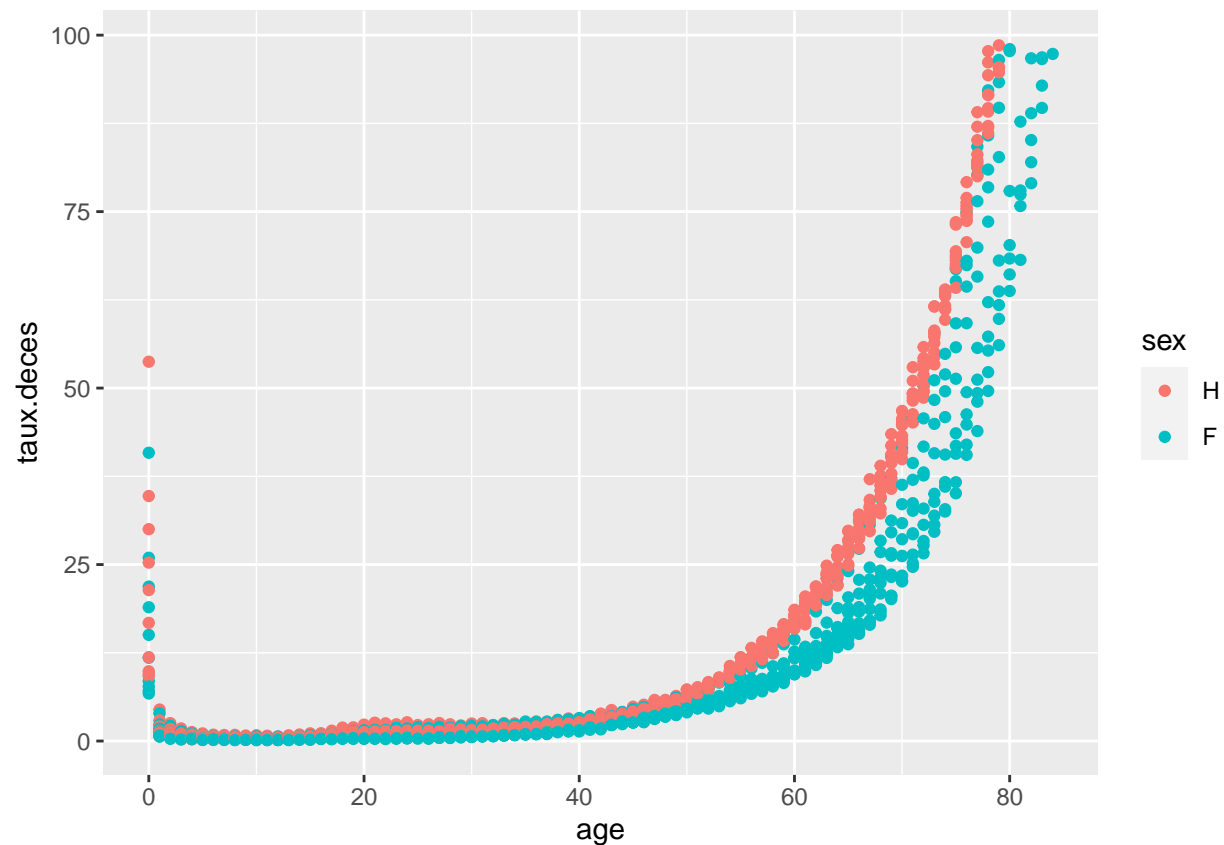
$$\text{taux.deces} = \frac{1000 \times \text{nb.deces}}{PA}$$

Les 15 variables d'identification du décès sont calculés de la même manière.

```
# summary(mort)
ggplot(mort, aes( y=taux.deces, fill=sex))+
  geom_boxplot(aes(x=periode))
```



```
ggplot(mort, aes( y=taux.deces, fill=sex, color=sex))+
  geom_point(aes(x=age))
```



TODO : description du data avec stat descriptives ?

```
head(mort)
```

```
##   i.. age periode sex      PA nb.deces taux.deces Infections Cancer
## 1  1  0  43-47  H 224654   12072    53.736    3.494  0.098
## 2  2  0  43-47  F 212730    8688    40.841    2.994  0.061
## 3  3  0  48-52  H 202692    7035    34.708    1.692  0.074
## 4  4  0  48-52  F 192602    5001    25.965    1.516  0.078
## 5  5  0  53-57  H 193164    5797    30.011    0.601  0.119
## 6  6  0  53-57  F 182400    3991    21.880    0.450  0.088
##   Tumeurs.Bénignes Endocrine Maladies.Sang Systeme.Nerveux Cerebrovasculaire
## 1          0.085      0.347         0.089          1.282             0.013
## 2          0.056      0.259         0.127          1.011             0.009
## 3          0.039      0.222         0.074          0.834             0.015
## 4          0.067      0.119         0.036          0.597             0.021
## 5          0.036      0.140         0.026          0.766             0.047
## 6          0.044      0.088         0.016          0.488             0.000
##   Cardiaque Maladies.Respiratoires Maladies.Foie Maladies.Digestives
## 1      0.036             10.826         0.027             5.092
## 2      0.033              8.739         0.038             3.685
## 3      0.049              6.276         0.054             1.919
## 4      0.026              4.875         0.042             1.231
## 5      0.047              3.018         0.031             1.066
## 6      0.055              2.489         0.033             0.691
##   Maladie.Génito_urinaire Non.Definis Mort.Naturelle Mort.Violente
```

```
## 1      0.125      1.340      29.962      0.921
## 2      0.132      1.020      21.976      0.700
## 3      0.049      0.745      21.718      0.947
## 4      0.062      0.633      15.919      0.742
## 5      0.031      0.564      22.810      0.709
## 6      0.027      0.378      16.541      0.493
```

Entre les années 1943 et 1948, 224 654 nourissons étaient de sexe masculins. 12 072 d'entre eux sont morts. Ce qui représente environ 53 individus issus de cette catégorie sur 1000 sont morts sur cette période. **TODO plus détailler une autre ligne**

```
mort <- mort %>% filter((age >=18) & (age <=25))
mort$age <- mort$age %>% factor()
```

Modèle a choisir, lequel on fait (beaucoup de variables, avec pleins de catégories)

Modalité de référence

```
#contrasts(mort$age)
contrasts(mort$periode)
```

```
##      48-52 53-57 58-62 63-67 68-72 73-77 78-82 83-87 88-92
## 43-47    0    0    0    0    0    0    0    0    0
## 48-52    1    0    0    0    0    0    0    0    0
## 53-57    0    1    0    0    0    0    0    0    0
## 58-62    0    0    1    0    0    0    0    0    0
## 63-67    0    0    0    1    0    0    0    0    0
## 68-72    0    0    0    0    1    0    0    0    0
## 73-77    0    0    0    0    0    1    0    0    0
## 78-82    0    0    0    0    0    0    1    0    0
## 83-87    0    0    0    0    0    0    0    1    0
## 88-92    0    0    0    0    0    0    0    0    1
```

```
contrasts(mort$sex)
```

```
##      F
## H 0
## F 1
```

La modalité de référence est l'âge 18ans. La modalité de référence pour la période temporelle est entre 1943 et 1947. Enfin la modalité de référence au niveau sexe est l'homme. **TODO changer la modalité de ref de l'âge**

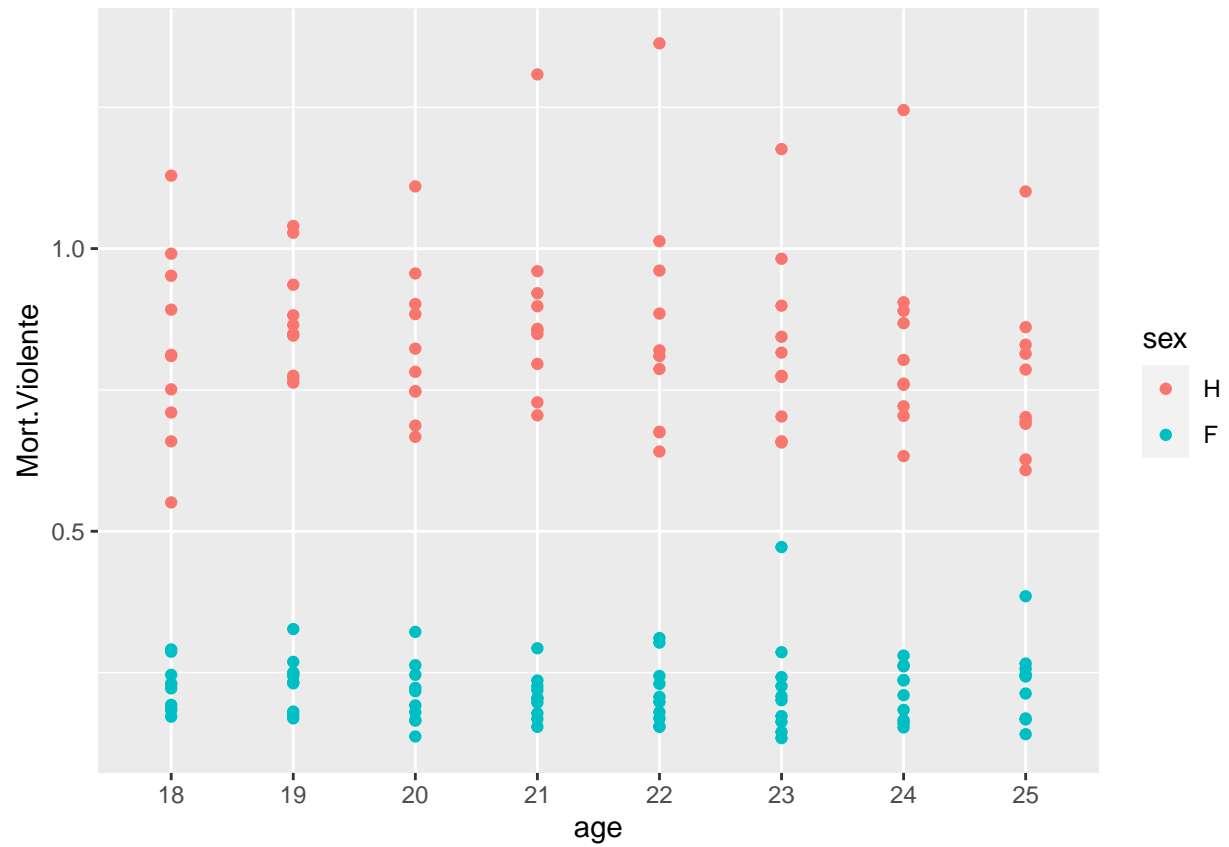
TODO choisir un modèle

```
# glm(nb.deces ~ periode * sex * age, offset = log(PA), data = mort)
# glm(Infections ~ periode * sex * age, offset = log(PA), data = mort)
# glm(Cancer ~ periode * sex * age, offset = log(PA), data = mort)
# glm(Tumeurs.Bénignes ~ periode * sex * age, offset = log(PA), data = mort)
#endocrine <- glm(Endocrine ~ periode + sex + age, offset = log(PA),
```

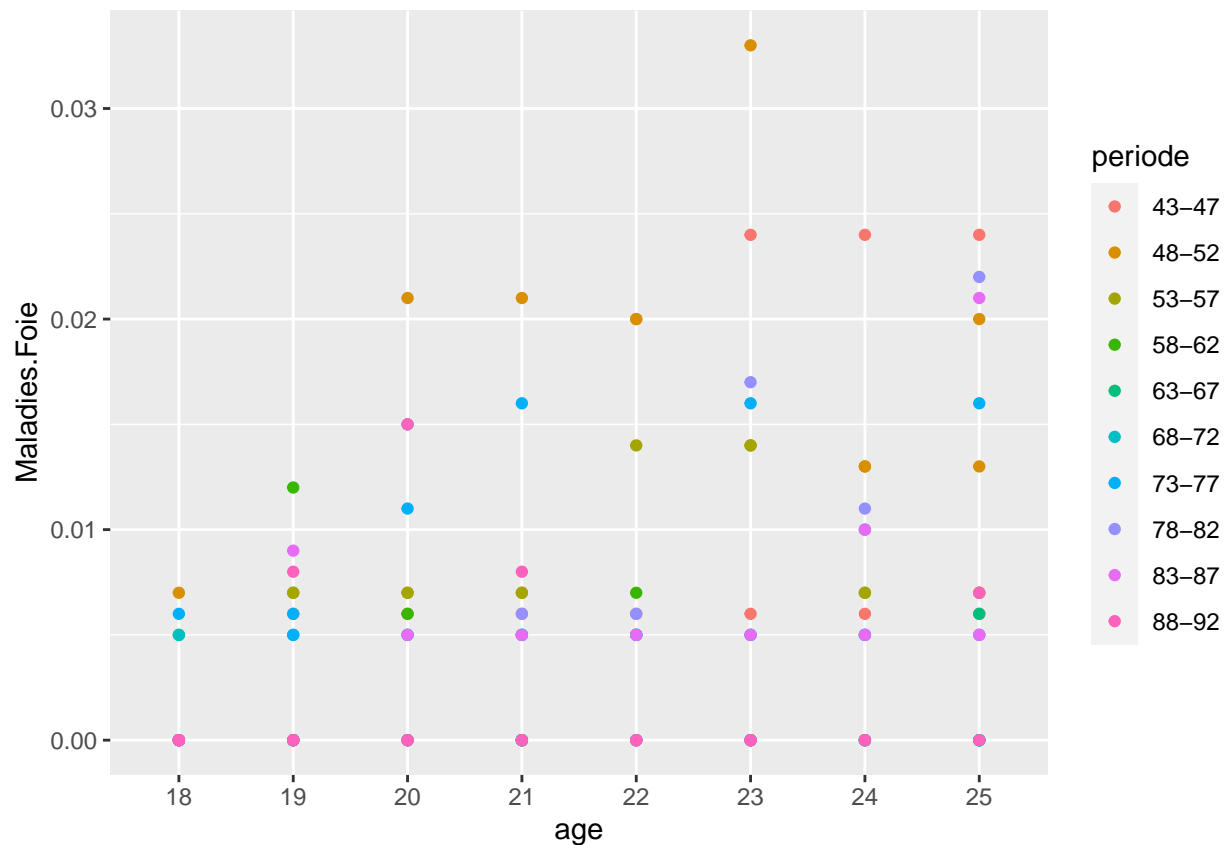
```
#           data = mort)
# glm(Maladies.Sang ~ periode * sex * age, offset = log(PA), data = mort)
# glm(Cerebrovasculaire ~ periode * sex * age, offset = log(PA), data = mort)
# glm(Cardiaque ~ periode * sex * age, offset = log(PA), data = mort)
# glm(Maladies.Respiratoires ~ periode * sex * age, offset = log(PA), data = mort)
# glm(Maladies.Foie ~ periode * sex * age, offset = log(PA), data = mort)
# glm(Maladies.Digestives ~ periode * sex * age, offset = log(PA), data = mort)
# glm(Maladie.Génito_urinaire ~ periode * sex * age, offset = log(PA), data = mort)
# glm(Non.Definis ~ periode * sex * age, offset = log(PA), data = mort)
# glm(Mort.Naturelle ~ periode * sex * age, offset = log(PA), data = mort)
glm(Mort.Violente ~ sex+periode+age, offset = log(PA), data = mort)
```

```
##
## Call: glm(formula = Mort.Violente ~ sex + periode + age, data = mort,
##          offset = log(PA))
##
## Coefficients:
## (Intercept)          sexF periode48-52 periode53-57 periode58-62
## -10.971139      -0.582776      -0.194360      -0.133210      -0.246109
## periode63-67 periode68-72 periode73-77 periode78-82 periode83-87
## -0.458640      -0.443070      -0.367310      -0.308913      -0.446977
## periode88-92          age19          age20          age21          age22
## -0.084656      0.034645      0.004651      0.032672      0.025000
##          age23          age24          age25
## 0.016780      0.019074      0.008079
##
## Degrees of Freedom: 159 Total (i.e. Null); 142 Residual
## Null Deviance: 19.1
## Residual Deviance: 1.65 AIC: -239.9
```

```
ggplot(mort)+
  geom_point(aes(x=age, y=Mort.Violente, color=sex))
```



```
ggplot(mort)+
  geom_point(aes(x=age, y= Maladies.Foie, color=periode))
```

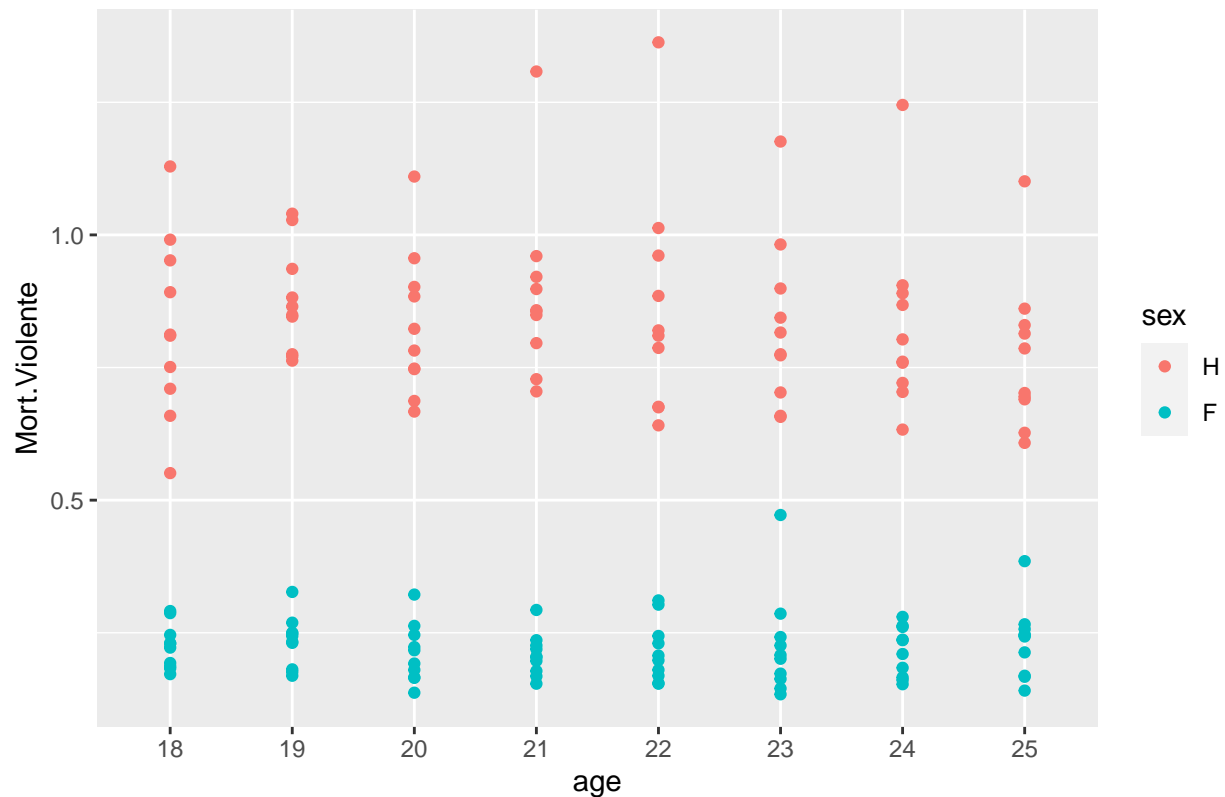


Etude des morts violentes chez les 18-25ans danois

Lorsqu'on parle de morts violentes, on fait souvent allusion à des décès par traumatisme ou empoisonnement accidentels, par suicide ou par homicide. Regardons tout d'abord un graphique en fonction du sexe de la personne.

```
ggplot(mort)+
  geom_point(aes(x=age, y=Mort.Violente, color=sex))+
  ggtitle("Morts violentes en fonction de l'âge")
```

Morts violentes en fonction de l'âge



Sur ce graphique, on voit clairement une différence entre les hommes et les femmes par rapport aux morts violentes.

```
model_nb_deces<-glm(nb.deces ~ periode + sex + age, offset = log(PA),
                    family = "poisson", data = mort)
#model_nb_deces
summary(model_nb_deces)
```

```
##
## Call:
## glm(formula = nb.deces ~ periode + sex + age, family = "poisson",
##      data = mort, offset = log(PA))
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -5.9342  -1.3113  -0.4069   1.0146   7.0480
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -5.98362    0.02279 -262.586 < 2e-16 ***
## periode48-52 -0.62843    0.02435  -25.806 < 2e-16 ***
## periode53-57 -0.81550    0.02618  -31.153 < 2e-16 ***
## periode58-62 -0.91484    0.02643  -34.620 < 2e-16 ***
## periode63-67 -1.00699    0.02541  -39.628 < 2e-16 ***
## periode68-72 -0.95409    0.02442  -39.072 < 2e-16 ***
## periode73-77 -0.95281    0.02499  -38.123 < 2e-16 ***
```



```

## periode78-82 -0.89548    0.02457   -36.442 < 2e-16 ***
## periode83-87 -1.04554    0.02534   -41.265 < 2e-16 ***
## periode88-92 -1.00979    0.02849   -35.444 < 2e-16 ***
## sexF          -0.77775    0.01369   -56.798 < 2e-16 ***
## age19         0.06000    0.02578    2.327 0.019944 *
## age20         0.05344    0.02584    2.069 0.038593 *
## age21         0.07829    0.02569    3.048 0.002307 **
## age22         0.08240    0.02567    3.210 0.001326 **
## age23         0.09315    0.02562    3.635 0.000278 ***
## age24         0.12540    0.02547    4.924 8.48e-07 ***
## age25         0.10770    0.02562    4.203 2.63e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##      Null deviance: 7210.33  on 159  degrees of freedom
## Residual deviance:  624.26  on 142  degrees of freedom
## AIC: 1738
##
## Number of Fisher Scoring iterations: 4

```

Parmi la strate 43-47, on fait la comparaison entre deux sous-strates :

- La sous-strate “âge 18ans et sex=H” -> le rapport $\frac{Mort.violentes}{pa}$ est $e^{-5.98}$

- La sous strate “âge 18ans et sex=F” -> le rapport $\frac{Mort.violentes}{pa}$ est $e^{-5.98}e^{-0.78}$

Si on compare les deux dernières sous-strate :

Le rapport d’incidence moyenne est entre le sexe masculin et le sexe féminin, à âge fixé à 18 ans est de $e^{-0.78} = 0.458$.